Tag and Go Seek: Ensuring Successful Tagging, Discoverability and Reusability of Content

Dr. Adelaide K. Cherry
Air Force Institute for Advanced Distributed Learning
Maxwell AFB—Gunter Annex, Alabama
Adelaide.Cherry@maxwell.af.mil

Dr. Thomas D. Wason
Teleologic Learning Company
Atlanta, Illinois
twason@teleologic.net

ABSTRACT

Locating educational courses in a distributed environment is a daunting task. This is made even more demanding when the courses have been subdivided into reusable modules, as the number of objects rapidly increases. Metadata—searchable descriptions of the objects—supports this process. This report, sponsored by the Air Force Institute for Advanced Distributed Learning and funded by the Joint ADL Co-Lab, describes the process and findings of a project to develop guidelines and taxonomies to facilitate content reuse. The project's goal was to develop an appropriate set of metadata fields and the vocabularies and taxonomies to be used to populate those fields. The project methodology featured a meeting of course developers at several levels collaborating to define metadata and the rules for reuse. There are five significant outcomes from this project: (1) A set of defined metadata fields; (2) Value domains including three existing sources for the primary taxonomies; (3) Models and rules for module development and reuse; (4) Validation of methodology; (5) Prototype tagging tool. The methods used during the face-to-face meeting of course developers may prove valuable to others in the development of metadata fields and taxonomies. Participants focused on real courses and modules; a concrete task to force fit a module into multiple courses, thus ensuring reusability, while multiple independently tasked recorders captured the process. The discovery of existing resources for subject, application domain and proficiency level leads to the possibility of existing sources of taxonomies for other career fields. Team agreement on a common course structure model to support reusable modules was a surprising result. The metadata fields, value domains, and course structures model are compatible with SCORM and attest to the utility of the SCORM specifications. The method and the project results need to be tested for replication, but offer a working model that can be used by the broader ADL community.

ABOUT THE AUTHORS

Dr. Adelaide K. Cherry currently serves as the Academic Advisor to the Commander of the Air Force Institute for Advanced Distributed Learning (AFIADL). She is the Institute's senior educational advisor and consultant for Air Force ADL implementation and oversees special studies and projects. Dr. Cherry spearheaded the establishment of the ADL Consulting Division responsible for managing the Air Force's Integrated Learning Center that includes an enterprise-level learning management system, object and media repository, and an electronic customer support center. She has lectured and conducted workshops in leadership and management, distance learning, curriculum development, educational technology, and other subjects. Dr. Cherry is the author of several publications and white papers including Brilliant Warrior: Information Technology Integration in Education and Training, for the futuristic Air Force 2025 study. She has a Ph.D. in Administration and Higher Education from the University of Alabama.

Dr. Thomas D. Wason is a Learning Architect at Teleologic Learning Company. He led the original IMS educational metadata effort and participated in the development of the Learning Object Metadata (LOM) specification of the IEEE Learning Standards Technology Committee. LOM is used in the Sharable Content Object Reference Model (SCORM). He has used the ADL's SCORM on a previous AFIADL project. Dr. Wason has worked on eight international standards for network and internet technologies. He was director of research and evaluation at the Institute for Academic Technology at UNC-Chapel Hill. He is a reviewer for the National Science Foundation on metadata and digital libraries. Dr. Wason earned his undergraduate degree in engineering from MIT and his MS and Ph.D. in experimental psychology from North Carolina State University.

INTRODUCTION

The Air Force must train its personnel. This is a continual, recurrent critical need that must address a large population. The Air Force expends considerable resources to develop effective training materials. This project to define metadata and taxonomies explores the potential for the reuse of instructional modules to achieve greater efficiencies and return on investment. Locating appropriate educational resources in a distributed environment is a daunting task as the number of objects rapidly increases. This task can be made easier if metadata—searchable descriptions of the objects—are congruent with the vocabularies of the user communities. The Air Force Institute for Advanced Distributed Learning (AFIADL), with funds provided by the Joint ADL Co-Lab, sponsored a project to develop standard metadata that could be used across content domains and integrated into the Air Force course development process. Boeing Learning Systems and Teleologic Learning Company undertook the project. The project was executed with the considerable participation of USAF personnel from several bases through the efforts of AFIADL.

There are five concrete results from this project that will contribute to the successful tagging, discoverability, and reusability of content:

1. A set of defined metadata fields applicable across three domains
2. Value domains that map to standard Air Force training documents, including three existing sources for the primary taxonomies
3. Models and rules for module development and reuse that will guide future content development efforts
4. A method for validation of outcomes that will inform future studies
5. Prototype tagging tool that will ensure uniform tagging of content

In addition to these intended outcomes, there are other benefits serendipitously derived from this project. First, the project developed a method for defining metadata fields and their taxonomies that is based on standard processes and documents currently in use in the Air Force. This will enable the method and the metadata fields to be more readily incorporated across multiple content domains. Also, the metadata fields identified during this process are interoperable with the metadata schema of AFIADL's object repository and the Sharable Content Object Reference Model (SCORM) metadata—a surprising result as the participants in this project had no knowledge of SCORM beforehand. Boeing made use of this interoperability to develop a metadata tagging system and integrate it into the ADL repository. Both the metadata development method and the interoperability with SCORM are powerful results.

METHODS

The major objective of the project was to develop a set of metadata fields describing online courses and modules consisting of parts of courses. Describing parts of online courses with searchable metadata is an important aspect of reuse, the golden goal of online course development. Reuse reduces costs. The metadata has two aspects: the fields and the values (vocabularies and taxonomies) for those fields. The values are from a "value domain." The productivity of searching will improve as the community of "taggers" and searchers use the same vocabularies. Toward this end, standard taxonomies are desired.

A concrete method was used for defining the metadata fields and taxonomies. A primary issue was the potential variance of terminology for common concepts across different Air Force career fields. The method was designed to leverage the expertise of users from different career fields who, independently of each other, had developed instruction on the same concept. To initiate the process, several courses with units on fundamentals of electronics were identified; from these, three were selected. A single module from the fundamentals of electronics was selected from one of the courses as a prototypical reusable instructional module (RIM). A number of subject matter expert (SME) course developers from different Air Force bases were recruited to participate in a Taxonomy Working Group (TWG). A criterion for selection was experience developing content for one of the three pre-selected courses. The purpose of the TWG was to analyze the common concept being taught—in this case, fundamentals of electronics—and reach consensus on a common language, or vocabulary, for describing the content module.

The participants represented different career fields and were each associated with one of three selected courses. None of the participants had previous experience with metadata. Materials describing the purpose of the TWG and of the working methods were prepared. Participants were supplied with these materials and descriptions of the concepts of reusable

instructional modules, metadata and taxonomies. The three target courses and the prototypical RIM were also provided.

The TWG, hosted by AFIADL, met for 2 days in January 2004 at Maxwell AFB-Gunter Annex in Montgomery, AL. The general method for developing information had three components:

1. Task: Focus on the main objective (reuse)
2. Content: Use of a concrete instance ("force fit" a RIM)
3. Data: Use independently tasked recorders

All decisions were made by consensus. Discussion continued until consensus was reached. In all cases, all issues were resolved successfully. The TWG SME participants were not familiar with either SCORM or LOM (Learning Object Metadata). No attempt was made to educate them about either. Metadata, reuse strategies, and modularization were developed without introduction of either. Although the members of the facilitating organization were aware of both SCORM and LOM and may have directed discussions accordingly, this was not intended.

Members of AFIADL selected courses, recruited participants, arranged meeting logistics, and attended the meeting. They participated by asking questions and explaining the needs, setting up the subsequent teleconference calls, and coordinating the work between Boeing and AFIADL's technical staff to create the tagging tool's technical interface. AFIADL worked closely with Teleologic in setting up the meeting. Teleologic designed the experimental protocol, developed participant education materials, led the meeting, and wrote meeting reports. Teleologic and Boeing collaboratively developed the requirements for reusability for the selection of the concrete examples used.

Task: Focus on the Main Objective (Reuse)

The TWG was specifically tasked to define metadata fields and taxonomies that would be useful in locating the target RIM. Participants approached the task from the standpoint of SME authors developing existing courses. Their objective as course developers was to "discover" the target RIM through the use of metadata and other strategies they felt appropriate. They were asked to develop the metadata fields that describe a RIM and to determine the vocabularies and taxonomies necessary.

Content: Use of a Concrete Instance ("Force Fit" a RIM)

The target RIM, fundamentals of electronics, was to be "force-fit" into each of the three courses. The TWG considered what adjustments would need to be made to the RIM to make it fit into each course. This addressed the problem of reusability of modules. The TWG also discussed what future adjustments would need to be made to course structure and the course development process to ensure modules are constructed for intentional reuse. The three career development courses (CDCs) selected were:

1. CDC 2A351, Avionic Systems
2. CDC 4A251A, Biomedical Equipment
3. CDC 2E151, Satellite/Wideband

Data: Use Independently Tasked Recorders

There were a number of tracks of information to be developed. In order to capture the information effectively, multiple recorders were used. Teleologic and Boeing supplied the recorders. Each recorder focused on one particular aspect of the information. During the meeting there were independently tasked recorders for:

- Reusability (Boeing)
- Metadata fields and value domains (Teleologic)
- Concepts (Teleologic)
- Terminology (Teleologic)

The recorder for metadata fields and value domains also led the meeting. Each recorder used data collection methods of his/her own devising, particular to the appropriate domain.

PROJECT RESULTS

The January 2004 meeting was successful. The participants all contributed actively, at times producing results that surprised the meeting's hosts. Subsequent work has refined and built on the results. An online forum was available for further discussions, but it was not used extensively. Included in the follow-on work was a conference call of all available TWG participants. The project has produced:

1. A set of defined metadata fields
2. Value domains including three existing sources for the primary taxonomies

3. Reusability guidelines
4. Method for validation of outcomes
5. A prototype tagging tool

A Set of Defined Metadata Fields

Thirty-five (35) metadata fields are currently identified. These are listed in the Appendix. Many of them will not be created through manual tagging. Some are intended primarily for document management activities. The TWG divided those fields not used for document and environment control functions into two general categories: search and information. Search fields are used as specific targets of searches, for example, *subject*, *skill level*, and *keywords*. Other fields are informational, to be returned, as a group or selectively, to provide additional information for resource selection. Such fields include *title*, *description*, technical *format*, *learning resource type*, and Career Field Education and Training Plan (*CFETP)*. Not all fields are mandatory.

The fields were subsequently mapped into the SCORM/LOM system. All fields were mapped successfully. The one area that is not completely resolved is the table of contents (TOC). The TOC is not normally metadata, but many participants felt it to be useful in the search process. Inspection of a module's TOC was felt to be potentially valuable.

There is concern over the fairly large number of fields identified. Some will be invisible to the user, of importance to the repository function. Other fields are a function of the manner in which the user's context either constructs the field or uses it to construct a search. The tagging tool effort has focused on those fields that the user is most apt to interact with directly.

The purposes of some fields are not immediately obvious. For example, *CFETP* is a standard Air Force document that serves as the "contract" under which a specific module was created. The CFETP specifies the subject content that must be mastered for each particular skill level with a specialty training standard (STS). Thus the CFETP provides the basis for two taxonomies that are familiar to all course developers in the Air Force: *subject* and *skill level*.

Value Domains Including Three Existing Sources for the Primary Taxonomies

The sources of the values for the metadata fields are the "value domains." The expectation is that terminology will differ among SMEs for the same concepts. For this reason, separate terminology and concept recorders were used. As they were expected to combine their work after the meeting, they sat together to do some coordination during the meeting. This proved unnecessary for this TWG. Subsequent conversations during the teleconference call indicated that such agreement might not always arise in others areas (e.g., business management); the use of terminology and concept recorders in such situations is encouraged.

All critical value domains were satisfied from existing sources within the Air Force. This was a pleasant surprise for all involved because it meant that familiar course development concepts and terms could be utilized as metadata values. Other value domains were readily agreed upon.

Three value domain taxonomies from existing sources were identified, satisfying four of the field value domains:

1. Subject
2. Application Domain
3. Skill Level
4. Teaching Objectives

Subject

The subject value taxonomy is derived from the plan of instructions (POIs) of several courses through harmonization. The participants did not feel such harmonization would present difficulties, particularly since each POI was created to satisfy the CFETP's Career or Specialty Training Standard (C/STS), which contains a hierarchy of required topics. It was determined that the POI would be the source of the taxonomy for the subject field. An example of a subject is electronic fundamentals.

Application Domain

The application domain refers to the context in which the subject is applied. For example, radio power supplies for F-15 aircraft—in this case the F-15 aircraft could be an application domain for electronic fundamentals. The Air Force Technical Order (TO) numbering system will provide all the domains and sub-domains for the required metadata and was considered by all participants to be a satisfactory taxonomy for this field.

The Biomedical occupations do not use the TO numbering system and would constitute a different application domain. The Air Force Medical Logistics Letter (AFMLL) was proposed for this purpose.

Skill Level

Skill level applies to each subject level within the C/STS and is part of an established proficiency code system in the Air Force. The proficiency code classification is a two-level taxonomy: knowledge and application. It relates to the level of learning for each objective and applies to the RIMs within a course and the course itself, as these map from the POI to the C/STS. For example, a "2b" proficiency code indicates "knowledge of the procedures and is partially proficient at performing" the operation check of a radio on an F-15 aircraft.

Teaching Objectives

The teaching objectives taxonomy is associated with the C/STS, and by extension, the POI. At each topical reference in the C/STS, it was determined that the teaching objective can be constructed as the topic reference taxonomy and one or two additional levels of taxonomy reflecting the skill level appropriate for each topic element.

Reusability Guidelines

Discussions on reusability were ongoing during the course of the meeting. Reusability was introduced very early in the meeting with a presentation from the Boeing representative. At issue was the instructional value of "generic" versus "context specific" modules. The TWG participants agreed that the fundamentals of electronics content has considerable potential for reuse within the three career fields that were studied. The electronics module was applicable as review material within level 5 courses (the current lowest level of online training of a sequence of levels of Air Force technical training: 3, 5, 7, 9), but would require some modification.

Courses would also require designs that are particularly conducive to the use of reusable modules. One of the objectives of course design is the consistent use of application domain-specific reference (context specific learning) throughout the course to maintain relevance for the learner. For example, if the course is focused on training for avionics specialties, references to avionics should be made throughout. This could be difficult to accomplish if reusable modules are used. The participants all agreed on a model for course structure. Application-specific modules would alternate with generic modules. These would be encapsulated within a larger module (Figure 1).

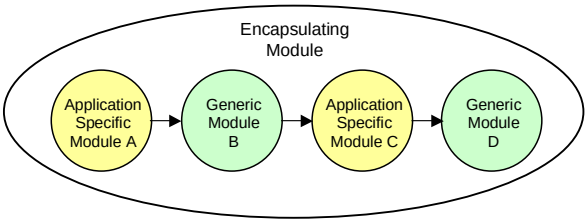

Figure 1. Model with alternating application-specific and generic modules

This model structure could be accomplished in either of two ways: (1) during an initial design, generic modules would be drawn from a library to intersperse between the custom application-specific modules and (2) an existing encapsulating module would be used, and the application-specific modules would be replaced as needed. This model of encapsulated modules made the use of the table of contents particularly attractive for course editing.

The acceptance of this model came as a surprise to the meeting hosts, including the instructional designers at AFIADL. This conclusion was tested at several points during the meeting, with consistent results.

Method for Validation of Outcomes

The intended method for validating the metadata and taxonomies is to have other authors/SMEs attempt to discover the target module on fundamentals of electronics. The authors will have the metadata fields present and will use the taxonomies. The validation method will analyze the efficiency of the metadata system developed by the TWG. This efficiency will be determined from the search successes of the testing SMEs. Changes in the metadata system will be proposed as appropriate.

At the time of this writing, the validation has not yet been performed. The tagging tool has been developed and integrated into the ADL repository and will serve as the method for testing the metadata fields instead of using paper methods.

Prototype Tagging Tool

A prototype metadata tagging and search tool is a product developed by Boeing Learning Systems to ensure reliable tagging by multiple users. It interfaces with the existing AFIADL metadata database. In developing the metadata tagging tool, consideration was given to methods of metadata tagging, particularly within the workflow. The future workflow process may include some tagging to be done automatically, some manually. The prototype tool in its current state

is designed for the manual tagging process. In the future, during searching, some metadata fields may infer the query's intent. Presentation type, such as graphical or verbal, may be inferred from the MIME type in the *Format* field. Similarly, the existence of assessment would be inferred from the contents of the *Learning Resource Type* (e.g., exercise, self-assessment, exam, problem statement).

The tagging tool used the SCORM/LOM realization of the metadata fields, communicating with the AFIADL metadata database in that format. This has proven largely effective.

CONCLUSIONS

The TWG method appears to have been successful. The use of multiple teams of SMEs in defining both metadata fields and taxonomies was effective. Approaches to course authoring and reuse strategies were developed. Existing sources of important taxonomies were revealed. The process was one of mutual education. For example, when one of the taxonomies (the TO numbering system for *application domain)* was "discovered," the leader commented that this was a "pleasant surprise." One of the participants responded that "We knew the answer; you had to ask the right question." All found this amusing and true. The use of a concrete task facilitated the work process while still producing broad-based results. The use of tasked recorders was useful where needed. The fact that working taxonomies were found to exist reduced the need for all recorders; however, they contributed substantially to the final report with their notes and observations. As the discovery of pre-existing useful taxonomies was not known at the outset, it would have been difficult (and probably foolhardy) to have omitted recorders. It appears that the method may have future usefulness. The quality of the result from the meeting was sufficient to support the development of a preliminary metadata tagging tool.

SCORM's LOM was shown to be effective in mapping the fields defined by the TWG to the database previously established by AFIADL. Without attempting to use SCORM as a foundation, it served as a basis for interoperability among the defined metadata tools, the metadata tagging and search tools, and the existing AFIADL database. The TWG metadata fields were mapped to the LOM. The metadata tagging tools then mapped the LOM to the database. As the database had been originally set up to use LOM, the mapping was successful. A number of the TWG's fields were realized using the LOM classification system. The general nature of this part of the LOM was both an asset—it allowed a *de facto* extension to LOM—and a limitation, as the AFIADL's database did not have explicit classifications that matched all TWG field classifications.

The proposed model for course structure for reusability was unexpected. It relies in part on the table of contents; hence the TOC was also considered to be a potential module access route. If a (sub)RIM were located at a node in a larger RIM's TOC, the participants felt browsing that (sub)RIM may facilitate discovery. TOC, although of interest as part of the process for course development and RIM importation, was considered outside the scope of this project. The TOC is more a constituent of the SCORM manifest within the <organization> portion, although the <item> nodes within the <organization> do contain metadata.

The general conclusion was that all tasks for this project were concluded successfully, demonstrating the effectiveness of the meeting plan and the project design. Follow-on studies to monitor the reliability of results as well as to develop guidelines and protocols for the Air Force are needed.

Appendix

Metadata Fields and SCORM Equivalents

* = Identified vocabulary or taxonomy
** = Air Force or DoD taxonomy

Field	SCORM/LOM
Identifier	General > Identifier
Title	General > Title
Catalog Entry	General > CatalogEntry
Description	General > Description
Keyword	General > Keyword
Version	Lifecycle > Version
Life Cycle Status *	Lifecycle > Status
Publisher	Lifecycle > Contributor > Role = Publisher
Date	Lifecycle > Contributor > Role = Publisher, Date = …
Contributor *	See Publisher.
POC	See Publisher. Possibly extend Contributor Role vocabulary.
Metametadata : Contribute	Metametadata > Contribute > Role = Publisher
Metametadata : Metadatascheme	Metametadata > MetadataScheme
Technical Format *	Technical > Format
System Requirements	Technical > Requirement > Type = …
Learning Resource Type: *	Educational > LearningResourceType
Contains Assessment *	Educational > LearningResourceType
Instructional Delivery Mode *	Educational > Context
Average Seat Time	Educational >TypicalLearningTime
Cost *	Rights > Cost
Copyright or Trademark *	Rights > Copyrightand OtherRestrictions
Content Source	Relation > Kind = IsBasedOn
Subject **	Classification > Purpose = Discipline
Domain **	Classification > Purpose = Domain
Application *	Classification > Purpose = Application
Skill Level **	Classification > Purpose = Skill Level. Alternative: Educational > Context.
Level in Table of Contents **	Classification > Purpose = TOCLevel
CFETP	Classification > Purpose = CFETP, Taxonpath(s).
Teaching Objectives **	Classification > Purpose = Educational Objectives Note: In addition to the Entry and the ID of the specific C/STS, each C/STS taxonpath terminal taxon has one or two nested taxons. The proficiency codes comprise a taxonomy that may entail one or two levels, depending on the specific code.
Prerequisites	Classification > Purpose = Prerequisites. Description with optional use of Taxonpath.
Modifiable *	Classification > Purpose = Modifiable
Convertible *	Classification > Purpose = Convertible
Tracking *	Classification > Purpose = Tracking Note: SCORM 1.3 will support tracking. The metadata system does not currently have a flag to indicate this capability.
Sensitivity Status **	Classification > Purpose = Security Level
Accessible *	Classification > Purpose = Accessibility Restrictions